Lab8A : OB Decomposition and IV in Stata

Introduction to Econometrics,Fall 2020

Yi Wang

Nanjing University

24/12/2020

Section 1

Oaxaca-Blinder Decomposition in Stata

Subsection 1

Review and Introduction

Oaxaca-Blinder Decomposition in Stata

- Review the Theory

Oaxaca-Blinder Decomposition

Oaxaca-Blinder Decomposition: difference in mean

- The difference in mean of Y_i of group A and B is

$$\bar{Y}_A - \bar{Y}_B = \hat{\beta}_A \bar{X}'_A - \hat{\beta}_B \bar{X}'_B$$

- A small trick: plus and minus a term $\hat{\beta}_B \bar{X}'_A$,then

$$\begin{aligned}\bar{Y}_A - \bar{Y}_B &= \hat{\beta}_A \bar{X}'_A - \hat{\beta}_B \bar{X}'_B \\ &= \hat{\beta}_A \bar{X}'_A - \hat{\beta}_B \bar{X}'_A + \hat{\beta}_B \bar{X}'_A - \hat{\beta}_B \bar{X}'_B \\ &= (\hat{\beta}_A - \hat{\beta}_B)\bar{X}'_A + \hat{\beta}_B(\bar{X}'_A - \bar{X}'_B)\end{aligned}$$

- Then the second term is **characteristics effect** which describes how much the difference of outcome, Y, in mean is due to differences in the levels of explanatory variables(characteristics).
- the first term is **coefficients effect** which describes how much the difference of outcome, Y, in mean is due to differences in the magnitude of regression coefficients.

Oaxaca-Blinder Decomposition in Stata

- Commands-*oaxaca*-

```
help oaxaca
findit oaxaca
ssc install oaxaca
```

Oaxaca-Blinder Decomposition in Stata

- Syntax

```
oaxaca depvar [indepvars] [if] [in] [weight] , by(groupvar) [options]
```

Oaxaca-Blinder Decomposition in Stata

- Options
  - by(groupvar)：指定分组变量
  - swap：交换组别
  - threefold：three-fold decomposition; the default.
  - weight(#) / pooled / omega：two-fold decomposition; 设置参照系数(nondiscriminatory coefficients).
    - ★ weight(#)—#是Group1相对于Group2的权重，如weight(1)，将Group1系数设置为参照系数.
    - ★ omega—using pooled model excluding groupvar.
    - ★ pooled—using pooled model including groupvar.
  - relax—do no stop on dropped coefficients/zero variances.
  - noisily—oaxaca first estimates two group-specific regression models and then performs the decomposition. noisily causes the group models' results to be displayed.

Subsection 2

An Example : Twofold decomposition

Oaxaca-Blinder Decomposition in Stata

- An Example
  - The standard application of the Blinder–Oaxaca Decomposition is to divide **Male-female average wage gap** into two parts**:**
    - ⋆ **Explained Part**: due to differences in the **levels** of explanatory variables
    - ⋆ such as schooling years,experience,tenure,industry,occupation,etc.
    - ⋆ **Unexplained Part**: due to differences in the **coefficients** to explanatory variables
    - ⋆ such as returns to schooling years, experience and tenure and premium in industry and occupation, etc.
  - An example using data from the Swiss Labor Market Survey 1998 (Jann 2003)
  - Reference: Jann B. The Blinder–Oaxaca decomposition for linear regression models[J]. The Stata Journal, 2008, 8(4): 453-479.

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - Different weight
  - Remember: Weight $\beta^* = W \times \beta_0 + (I - W)\beta_1$

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - weight=1: $\beta^* = \beta_0$: Wage of men is viewed as non-discriminatory.

```
. oaxaca lnwage educ exper tenure, by(female) weight(1)

Blinder-Oaxaca decomposition                    Number of obs     =      1,434
                                                  Model           =     linear
Group 1: female = 0                               N of obs 1      =        751
Group 2: female = 1                               N of obs 2      =        683
```

lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	3.440222	.0174874	196.73	0.000	3.405947	3.474497
group_2	3.266761	.0218522	149.49	0.000	3.223932	3.309591
difference	.1734607	.027988	6.20	0.000	.1186052	.2283163
explained	.0908978	.0136196	6.67	0.000	.0642038	.1175917
unexplained	.082563	.0255804	3.23	0.001	.0324263	.1326996
explained						
educ	.047771	.0110011	4.34	0.000	.0262092	.0693328
exper	.0190709	.0060299	3.16	0.002	.0072526	.0308892
tenure	.0240559	.0064492	3.73	0.000	.0114156	.0366961
unexplained						
educ	-.0640559	.1193498	-0.54	0.591	-.2979772	.1698654
exper	-.0397237	.04058	-0.98	0.328	-.1192591	.0398117
tenure	.0420986	.0270201	1.56	0.119	-.0108598	.0950571
_cons	.1442439	.1340957	1.08	0.282	-.1185788	.4070667

```
. est store oa1
```

Oaxaca-Blinder Decomposition in Stata

- Threefold decomposition
  - **The decomposition output reports in the first panel** :
    - ⋆ 3.44—the mean of log wages (lnwage) for men
    - ⋆ 3.27—the mean of log wages (lnwage) for women
    - ⋆ 0.17—wage gap
    - ⋆ explained—the mean increase in women's wages if they had the same characteristics as men
    - ⋆ unexplained—the change in women's wages when applying the men's coefficients to the women's characteristics

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - weight=0: $\beta^* = \beta_1$: Wage of women is viewed as non-discriminatory.

```
. oaxaca lnwage educ exper tenure, by(female) weight(0)

Blinder-Oaxaca decomposition                    Number of obs     =      1,434
                                                  Model           =     linear
Group 1: female = 0                               N of obs 1      =        751
Group 2: female = 1                               N of obs 2      =        683
```

lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	3.440222	.0174874	196.73	0.000	3.405947	3.474497
group_2	3.266761	.0218522	149.49	0.000	3.223932	3.309591
difference	.1734607	.027988	6.20	0.000	.1186052	.2283163
explained	.0852798	.015693	5.43	0.000	.0545222	.1160375
unexplained	.0881809	.026692	3.30	0.001	.0358655	.1404963
explained						
educ	.0510912	.012239	4.17	0.000	.0271031	.0750792
exper	.0254173	.0088089	2.89	0.004	.0081522	.0426824
tenure	.0087714	.0086201	1.02	0.309	-.0081238	.0256665
unexplained						
educ	-.0673761	.1255358	-0.54	0.591	-.3134218	.1786696
exper	-.0460701	.0470666	-0.98	0.328	-.1383189	.0461788
tenure	.0573831	.0368225	1.56	0.119	-.0147877	.129554
_cons	.1442439	.1340957	1.08	0.282	-.1185788	.4070667

```
. est store oa2
```

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - weight=0.5: Reimers(1983)

```
. oaxaca lnwage educ exper tenure, by(female) weight(0.5)

Blinder-Oaxaca decomposition                    Number of obs     =      1,434
                                                  Model           =     linear
Group 1: female = 0                               N of obs 1      =        751
Group 2: female = 1                               N of obs 2      =        683
```

lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	3.440222	.0174874	196.73	0.000	3.405947	3.474497
group_2	3.266761	.0218522	149.49	0.000	3.223932	3.309591
difference	.1734607	.027988	6.20	0.000	.1186052	.2283163
explained	.0880888	.0136315	6.46	0.000	.0613715	.1148061
unexplained	.0853719	.0255607	3.34	0.001	.035274	.1354699
explained						
educ	.0494311	.0112121	4.41	0.000	.0274558	.0714064
exper	.0222441	.0067646	3.29	0.001	.0089858	.0355024
tenure	.0164136	.0056741	2.89	0.004	.0052925	.0275347
unexplained						
educ	-.065716	.1224423	-0.54	0.591	-.3056985	.1742665
exper	-.0428969	.0438153	-0.98	0.328	-.1287734	.0429796
tenure	.0497409	.0318942	1.56	0.119	-.0127706	.1122523
_cons	.1442439	.1340957	1.08	0.282	-.1185788	.4070667

```
. est store oa3
```

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - weight the coefficients by the group sizes: Cotton(1988)

```
. sum female
. local p_1=1-r(mean)

. oaxaca lnwage educ exper tenure, by(female) weight(`p_1´)

Blinder-Oaxaca decomposition                    Number of obs     =      1,434
                                                  Model           =     linear
Group 1: female = 0                               N of obs 1      =        751
Group 2: female = 1                               N of obs 2      =        683
```

lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	3.440222	.0174874	196.73	0.000	3.405947	3.474497
group_2	3.266761	.0218522	149.49	0.000	3.223932	3.309591
difference	.1734607	.027988	6.20	0.000	.1186052	.2283163
explained	.0878688	.0137253	6.40	0.000	.0609678	.1147698
unexplained	.085592	.0256087	3.34	0.001	.0353997	.1357842
explained						
educ	.0495611	.0112649	4.40	0.000	.0274824	.0716398
exper	.0224927	.0068879	3.27	0.001	.0089927	.0359927
tenure	.015815	.0057996	2.73	0.006	.0044481	.027182
unexplained						
educ	-.065846	.1226846	-0.54	0.591	-.3063033	.1746113
exper	-.0431454	.0440695	-0.98	0.328	-.12952	.0432291
tenure	.0503395	.0322786	1.56	0.119	-.0129253	.1136043
_cons	.1442439	.1340957	1.08	0.282	-.1185788	.4070667

```
. est store oa4
```

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - Omega: Neumark (1988), Oaxaca and Ransom (1994)

```
. oaxaca lnwage educ exper tenure, by(female) omega

Blinder-Oaxaca decomposition                    Number of obs     =      1,434
                                                  Model           =     linear
Group 1: female = 0                               N of obs 1      =        751
Group 2: female = 1                               N of obs 2      =        683
```

lnwage	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	3.440222	.0174586	197.05	0.000	3.406004	3.47444
group_2	3.266761	.0218042	149.82	0.000	3.224026	3.309497
difference	.1734607	.0279325	6.21	0.000	.118714	.2282075
explained	.0925597	.0140752	6.58	0.000	.0649728	.1201466
unexplained	.080901	.0243589	3.32	0.001	.0331585	.1286435
explained						
educ	.0506419	.0115857	4.37	0.000	.0279344	.0733495
exper	.021852	.0064912	3.37	0.001	.0091295	.0345744
tenure	.0200658	.0054145	3.71	0.000	.0094536	.030678
unexplained						
educ	-.0669269	.1395872	-0.48	0.632	-.3405128	.206659
exper	-.0425047	.0412027	-1.03	0.302	-.1232605	.038251
tenure	.0460887	.0271554	1.70	0.090	-.0071349	.0993122
_cons	.1442439	.1624352	0.89	0.375	-.1741233	.4626112

```
. est store oa5
```

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - Pooled: Jann(2008) and Fortin(2011)

```
. oaxaca lnwage educ exper tenure, by(female) pooled

Blinder-Oaxaca decomposition                    Number of obs     =      1,434
                                                  Model           =     linear
Group 1: female = 0                               N of obs 1      =        751
Group 2: female = 1                               N of obs 2      =        683
```

lnwage	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	3.440222	.0174586	197.05	0.000	3.406004	3.47444
group_2	3.266761	.0218042	149.82	0.000	3.224026	3.309497
difference	.1734607	.0279325	6.21	0.000	.118714	.2282075
explained	.089347	.0137531	6.50	0.000	.0623915	.1163026
unexplained	.0841137	.025333	3.32	0.001	.034462	.1337654
explained						
educ	.0493404	.0113168	4.36	0.000	.0271599	.071521
exper	.0215214	.0064081	3.36	0.001	.0089617	.0340811
tenure	.0184852	.0051833	3.57	0.000	.0083262	.0286443
unexplained						
educ	-.0656254	.139432	-0.47	0.638	-.3389072	.2076564
exper	-.0421741	.0411638	-1.02	0.306	-.1228537	.0385055
tenure	.0476693	.0271699	1.75	0.079	-.0055828	.1009213
_cons	.1442439	.1624352	0.89	0.375	-.1741233	.4626112

```
. est store oa6
```

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - Oaxaca-Blinder Table:

```
. esttab oa1 oa2 oa5 oa6,                          ///
        b(%6.3f) se(%6.3f)                          ///
        star(* 0.1 ** 0.05 *** 0.01) replace       ///
        mtitle(male female omega pooled)           ///
        obslast nogaps compress

. esttab oa1 oa2 oa5 oa6 using OB.csv,             ///
        b(%6.3f) se(%6.3f) r2(%6.3f)               ///
        star(* 0.1 ** 0.05 *** 0.01) replace       ///
        mtitle(male female omega pooled)           ///
        obslast nogaps compress
```

Oaxaca-Blinder Decomposition in Stata

- Twofold decomposition
  - Oaxaca-Blinder Table:

	(1) male	(2) female	(3) omega	(4) pooled
overall				
group_1	3.440***	3.440***	3.440***	3.440***
	(0.017)	(0.017)	(0.017)	(0.017)
group_2	3.267***	3.267***	3.267***	3.267***
	(0.022)	(0.022)	(0.022)	(0.022)
difference	0.173***	0.173***	0.173***	0.173***
	(0.028)	(0.028)	(0.028)	(0.028)
explained	0.091***	0.085***	0.093***	0.089***
	(0.014)	(0.016)	(0.014)	(0.014)
unexplai~d	0.083***	0.088***	0.081***	0.084***
	(0.026)	(0.027)	(0.024)	(0.025)
explained				
educ	0.048***	0.051***	0.051***	0.049***
	(0.011)	(0.012)	(0.012)	(0.011)
exper	0.019***	0.025***	0.022***	0.022***
	(0.006)	(0.009)	(0.006)	(0.006)
tenure	0.024***	0.009	0.020***	0.018***
	(0.006)	(0.009)	(0.005)	(0.005)
unexplai~d				
educ	-0.064	-0.067	-0.067	-0.066
	(0.119)	(0.126)	(0.140)	(0.139)
exper	-0.040	-0.046	-0.043	-0.042
	(0.041)	(0.047)	(0.041)	(0.041)
tenure	0.042	0.057	0.046*	0.048*
	(0.027)	(0.037)	(0.027)	(0.027)
_cons	0.144	0.144	0.144	0.144
	(0.134)	(0.134)	(0.162)	(0.162)
N	1434	1434	1434	1434

Standard errors in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Subsection 3

An Example : Robust and Bootstrap S.E.

Oaxaca-Blinder Decomposition in Stata

- Standard error: Robust and Bootstrap S.E.
  - bootstrap 自助法（又称为自举法）：是对原始样本进行“再抽样”的方法
  - 最常见实现自助法的方法：使用可选项 vce(bootstrap)
  - 例如：reg y x1 x2 x3, vce(boot,reps(400))
    reps(400)表明抽样的样本个数为400次
  - 对于一般的统计量使用自助法，可以使用命令bootstrap,
    如：bootstrap, _b _se, reps(400): reg y x1 x2 x3

Oaxaca-Blinder Decomposition in Stata

- Standard error: Robust and Bootstrap S.E.

```
qui oaxaca lnwage educ exper tenure, by(female) pooled
est store oase1

qui oaxaca lnwage educ exper tenure, by(female) pooled vce(robust)
est store oase2

qui oaxaca lnwage educ exper tenure, by(female) pooled vce(boot,r(
est store oase3

esttab oase1 oase2 oase3                    ///
   using OBSE.csv,                          ///
   b(%6.3f) se(%6.3f)                       ///
   star(* 0.1 ** 0.05 *** 0.01) replace     ///
   obslast nogaps compress
```

Subsection 4

An Example : Detailed OB decomposition

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition
  - Too long table: detail option

```
. tab isco, nofreq gen(isco)

. oaxaca lnwage educ exper tenure isco2-isco9, by(female) pooled detail
Blinder-Oaxaca decomposition                  Number of obs     =      1,434
                                                Model           =     linear
Group 1: female = 0                             N of obs 1      =        751
Group 2: female = 1                             N of obs 2      =        683
```

lnwage	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	3.440222	.0174589	197.05	0.000	3.406003	3.474441
group_2	3.266761	.0218047	149.82	0.000	3.224025	3.309498
difference	.1734607	.0279331	6.21	0.000	.118713	.2282085
explained	.0738838	.017772	4.16	0.000	.0390513	.1087163
unexplained	.0995769	.0266887	3.73	0.000	.047268	.1518859
explained						
educ	.0395615	.0097334	4.06	0.000	.0204843	.0586387
exper	.0218791	.0064781	3.38	0.001	.0091823	.0345759
tenure	.0180525	.0049008	3.68	0.000	.0084471	.0276579
isco2	-.0073025	.0059345	-1.23	0.218	-.0189339	.0043288
isco3	.0079376	.0059685	1.33	0.184	-.0037604	.0196357
isco4	.0148991	.0093095	1.60	0.110	-.0033473	.0331455
isco5	.0342628	.0101794	3.37	0.001	.0143115	.0542142
isco6	-.0067029	.0034618	-1.94	0.053	-.0134879	.0000821
isco7	-.0458051	.0114057	-4.02	0.000	-.0681599	-.0234503
isco8	-.0107627	.0044138	-2.44	0.015	-.0194135	-.0021119
isco9	.0078644	.0039715	1.98	0.048	.0000803	.0156485
unexplained						
educ	-.1324971	.1788045	-0.74	0.459	-.4829475	.2179533
exper	-.0345881	.0400924	-0.86	0.388	-.1131677	.0439914
tenure	.0475836	.0262981	1.81	0.070	-.0039597	.099127
isco2	-.0085307	.0228235	-0.37	0.709	-.0532641	.0362026
isco3	-.0461995	.0553921	-0.83	0.404	-.1547659	.062367
isco4	-.0489104	.0325967	-1.50	0.133	-.1127987	.0149779
isco5	-.002608	.0251929	-0.10	0.918	-.0519853	.0467692

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition
  - Too long table: detail option
  - detail：不加任何参数可以汇报各因素的单独贡献
  - detail()：在括号中定义因素组，聚合结果。
  - categorical()：识别dummy-variable sets，变换系数，使类别因素的分解结果不取决于基础组的选择。

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
. oaxaca lnwage educ exper tenure isco2-isco9, by(female) pooled  ///
     detail(exp_ten: exper tenure, isco: isco?) categorical(isco?)
(normalized: isco1 isco2 isco3 isco4 isco5 isco6 isco7 isco8 isco9)
Blinder-Oaxaca decomposition                    Number of obs     =      1,434
                                                  Model           =     linear
Group 1: female = 0                               N of obs 1      =        751
Group 2: female = 1                               N of obs 2      =        683
```

lnwage	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
overall						
group_1	3.440222	.0174589	197.05	0.000	3.406003	3.474441
group_2	3.266761	.0218047	149.82	0.000	3.224025	3.309498
difference	.1734607	.0279331	6.21	0.000	.118713	.2282085
explained	.0738838	.017772	4.16	0.000	.0390513	.1087163
unexplained	.0995769	.0266887	3.73	0.000	.047268	.1518859
explained						
educ	.0395615	.0097334	4.06	0.000	.0204843	.0586387
exp_ten	.0399316	.0089081	4.48	0.000	.022472	.0573911
isco	-.0056093	.012445	-0.45	0.652	-.0300009	.0187824
unexplained						
educ	-.1324971	.1788045	-0.74	0.459	-.4829475	.2179533
exp_ten	.0129955	.0400811	0.32	0.746	-.0655619	.0915529
isco	-.0159367	.0296549	-0.54	0.591	-.0740592	.0421858
_cons	.2350152	.195018	1.21	0.228	-.1472132	.6172435

```
exp_ten: exper tenure
isco: isco1 isco2 isco3 isco4 isco5 isco6 isco7 isco8 isco9
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
. ereturn display,coeflegend
```

lnwage	Coef.	Legend
overall		
group_1	3.440222	_b[overall:group_1]
group_2	3.266761	_b[overall:group_2]
difference	.1734607	_b[overall:difference]
explained	.0738838	_b[overall:explained]
unexplained	.0995769	_b[overall:unexplained]
explained		
educ	.0395615	_b[explained:educ]
exp_ten	.0399316	_b[explained:exp_ten]
isco	-.0056093	_b[explained:isco]
unexplained		
educ	-.1324971	_b[unexplained:educ]
exp_ten	.0129955	_b[unexplained:exp_ten]
isco	-.0159367	_b[unexplained:isco]
_cons	.2350152	_b[unexplained:_cons]

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
nlcom (educ_explained    : _b[explained:educ])          ///
      (expten_explained  : _b[explained:exp_ten])       ///
      (isco_explained    : _b[explained:isco])          ///
      (Total_explained   : _b[overall:explained])       ///
      (educ_unexplained  : _b[unexplained:educ])        ///
      (expten_unexplained: _b[unexplained:exp_ten])     ///
      (isco_unexplained  : _b[unexplained:isco])        ///
      (cons_unexplained  : _b[unexplained:isco])        ///
      (Total_unexplained : _b[overall:unexplained]), post
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
educ_expla~d:  _b[explained:educ]
expten_exp~d:  _b[explained:exp_ten]
isco_expla~d:  _b[explained:isco]
Total_expl~d:  _b[overall:explained]
educ_unexp~d:  _b[unexplained:educ]
expten_une~d:  _b[unexplained:exp_ten]
isco_unexp~d:  _b[unexplained:isco]
cons_unexp~d:  _b[unexplained:isco]
Total_unex~d:  _b[overall:unexplained]
```

lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
educ_explained	.0395615	.0097334	4.06	0.000	.0204843	.0586387
expten_explained	.0399316	.0089081	4.48	0.000	.022472	.0573911
isco_explained	-.0056093	.012445	-0.45	0.652	-.0300009	.0187824
Total_explained	.0738838	.017772	4.16	0.000	.0390513	.1087163
educ_unexplained	-.1324971	.1788045	-0.74	0.459	-.4829475	.2179533
expten_unexplained	.0129955	.0400811	0.32	0.746	-.0655619	.0915529
isco_unexplained	-.0159367	.0296549	-0.54	0.591	-.0740592	.0421858
cons_unexplained	-.0159367	.0296549	-0.54	0.591	-.0740592	.0421858
Total_unexplained	.0995769	.0266887	3.73	0.000	.047268	.1518859

```
. est store deoa1
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
. oaxaca lnwage educ exper tenure isco2-isco9, by(female) pooled                    ///
     detail(exp_ten: exper tenure, isco: isco?) categorical(isco?)

. nlcom (educ_explained    : _b[explained:educ]      / _b[overall:difference]*100)  ///
        (expten_explained  : _b[explained:exp_ten]   / _b[overall:difference]*100)  ///
        (isco_explained    : _b[explained:isco]      / _b[overall:difference]*100)  ///
        (Total_explained   : _b[overall:explained]   / _b[overall:difference]*100)  ///
        (educ_unexplained  : _b[unexplained:educ]    / _b[overall:difference]*100)  ///
        (expten_unexplained: _b[unexplained:exp_ten] / _b[overall:difference]*100)  ///
        (isco_unexplained  : _b[unexplained:isco]    / _b[overall:difference]*100)  ///
        (cons_unexplained  : _b[unexplained:_cons]   / _b[overall:difference]*100)  ///
        (Total_unexplained : _b[overall:unexplained] / _b[overall:difference]*100) , post
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
educ_expla~d:  _b[explained:educ]      / _b[overall:difference]*100
expten_exp~d:  _b[explained:exp_ten]   / _b[overall:difference]*100
isco_expla~d:  _b[explained:isco]      / _b[overall:difference]*100
Total_expl~d:  _b[overall:explained]   / _b[overall:difference]*100
educ_unexp~d:  _b[unexplained:educ]    / _b[overall:difference]*100
expten_une~d:  _b[unexplained:exp_ten] / _b[overall:difference]*100
isco_unexp~d:  _b[unexplained:isco]    / _b[overall:difference]*100
cons_unexp~d:  _b[unexplained:_cons]   / _b[overall:difference]*100
Total_unex~d:  _b[overall:unexplained] / _b[overall:difference]*100
```

lnwage	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
educ_explained	22.80716	5.60657	4.07	0.000	11.81849	33.79584
expten_explained	23.02053	5.683341	4.05	0.000	11.88139	34.15967
isco_explained	-3.233732	7.269927	-0.44	0.656	-17.48253	11.01506
Total_explained	42.59396	9.883593	4.31	0.000	23.22247	61.96545
educ_unexplained	-76.38447	102.0626	-0.75	0.454	-276.4234	123.6545
expten_unexplained	7.491894	23.41997	0.32	0.749	-38.41041	53.3942
isco_unexplained	-9.187481	17.16109	-0.54	0.592	-42.8226	24.44763
cons_unexplained	135.4861	108.2158	1.25	0.211	-76.61298	347.5852
Total_unexplained	57.40604	9.883593	5.81	0.000	38.03455	76.77753

```
. est store deoa2
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

```
esttab deoa1 deoa2,                    ///
     b(%6.3f) t(%6.3f)                 ///
     nogaps compress  obslast          ///
     star(* 0.1 ** 0.05 *** 0.01)      ///
     mtitle(coef percent(%))           ///
```

Oaxaca-Blinder Decomposition in Stata

- Detailed OB decomposition

	(1) coef	(2) percent_)
educ_exp_d	0.040***	22.807***
	(4.064)	(4.068)
expten_e_d	0.040***	23.021***
	(4.483)	(4.051)
isco_exp_d	-0.006	-3.234
	(-0.451)	(-0.445)
Total_ex_d	0.074***	42.594***
	(4.157)	(4.310)
educ_une_d	-0.132	-76.384
	(-0.741)	(-0.748)
expten_u_d	0.013	7.492
	(0.324)	(0.320)
isco_une_d	-0.016	-9.187
	(-0.537)	(-0.535)
cons_une_d	-0.016	135.486
	(-0.537)	(1.252)
Total_un_d	0.100***	57.406***
	(3.731)	(5.808)
N	1434	1434

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Section 2

IV estimator in Stata

Subsection 1

Review the Theory

IV estimator in Stata

- Review the Theory

Threatens to Internal Validity

- Three endogenous in OLS regression are:
  - **Omitted Variable Bias**(a variable that is correlated with X but is unobserved)
  - **Simultaneity or reverse causality Bias** (X causes Y,Y causes X)
  - **Errors-in-Variables Bias** (X is measured with error)
- One easy way to deal with these endogeneity is using **Instrumental Variable** method.

IV estimator in Stata

- Review the Theory

Instrumental Variable Method

Instrumental variables: 1 endogenous regressor & 1 instrument

- suppose a simple OLS regression like previous equation

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Because $E[u_i|X_i] \neq 0$, then we can use an instrumental variable(Z_i) to obtain an consistent estimate of coefficient.
- Intuitively, we want to split X_i into two parts:
  1. part that is correlated with the error term.
  2. part that is uncorrelated with the error term.
- If we can isolate the variation in X_i that is uncorrelated with u_i,then we can use this part to obtain a consistent estimate of the causal effect of X_i on Y_i.

IV estimator in Stata

- Review the Theory

Instrumental Variable Method

Instrumental variables: 1 endogenous regressor & 1 instrument

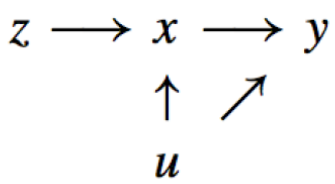
- An instrumental variable Z_i must satisfy the following 2 properties:
  1. **Instrumental relevance**: Z_i should be **correlated** with the casual variable of interest, X_i (endogenous variable),thus

$$Cov(X_i, Z_i) \neq 0$$

.

  2. **Instumental exogeneity**: Z_i is as good as randomly assigned and Z_i only affect on Y_i through X_i affecting Y_i channel.

$$Cov(Z_i, u_i) = 0$$



IV estimator in Stata

- Review the Theory

Instrumental Variable Method

IV estimator: Two Steps Least Square (2SLS)

- We can estimate the causal effect of X_i on Y_i in two steps

1. **First stage**: Regress X_i on Z_i & obtain predicted values of $\hat{X}_i$,if $Cov(Z_i, u_i) = 0$, then $\hat{X}_i$ contains variation in X_i that is uncorrelated with u_i

$$\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$$

.

2. **Second stage**: Regress Y_i on $\hat{X}_i$ to obtain the Two Stage Least Squares estimator $\hat{\beta}_{2SLS}$

$$\hat{\beta}_{2SLS} = \frac{\sum(Y_i - \bar{Y})(\hat{X}_i - \overline{\hat{X}})}{\sum(\hat{X}_i - \overline{\hat{X}})^2}$$

IV estimator in Stata

- Review the Theory

IV with Heterogeneous Causal Effects

Instrument Variables:Constant-effect

- Instrumental Variable is a useful method to make causal inference. It can eliminate
  - Omitted Variable Bias
  - Measurement Error
  - Reverse Causality
- Two Assumptions
  - Relevance(Weak Instrument): It can be test by the first stage regression and F-statistic.
  - Exogeneity: Can't be test formally but argue it using professional knowledges.
- Estimation and Inference
  - When IV satisify these two assumptions,the causal effect of coefficients of interest,TSLS estimator,β_{TSLS} can be NOT unbiased but **consistent**.
  - The sampling distribution of the TSLS estimator is also normal in large samples,so the general procedures for statistical inference in OLS can be used.

IV estimator in Stata

- Review the Theory

Some Practical Guides by Angrist and Pischke(2012)

Practical Guides

1. Check IV relevance
   - Always report the first stage and think about whether it makes sense(Signs and magnitudes)
   - Always report the F-statistic on the excluded instruments. The bigger,the better. Don't forget the rule of thumb.($F > 10$)
2. Check exclusion restriction
   - The exclusion restriction cannot be tested directly, but it can be falsified
   - Run and examine the reduced form(regression of dependent variable on instruments) and look at the coefficients, t-statistics and F-statistics for excluded instruments.
   - Because the reduced form is proportional to the casual effect of interest and is unbiased(OLS), so we should see the causal relation in the reduced form.If you can't see the causal relation in the reduced form,it's probably not there

IV estimator in Stata

- Review the Theory

Some Practical Guides by Angrist and Pischke(2012)

Practical Guides

3. Provide a substantive explanation for observed difference between 2SLS and OLS
   - How bid is the difference? What does this tell you?
   - Is the coefficient bigger when theory of endogeneity suggests it should be smaller? If so, why?
   - Measurement Error or heterogeneous effects?
4. If you have multiple instruments, report over-identification tests.
   - Pick your best single instrument and report just-identified estimates using this one only because just-identified IV is relatively unlikely to be subject to a weak bias.
   - Worry if it is substantially different from what you get using multiple instruments.
   - Check over-identified 2SLS estimates with LIML. LIML is less than precise than 2SIS but also less biased. If the results come out similar, be happy. If not, worry, and try to find stronger instruments.

Subsection 2

Introduction

IV estimator in Stata

- Syntax

```
ivregress estimator depvar [varlist1] (varlist2 = varlist_iv)
[if] [in] [weight] [, options]

help ivregress
```

IV estimator in Stata

- Options
  - estimator : 2sls/liml/gmm
  - depvar : dependent variable
  - varlist1 : exogenous variables
  - varlist2 : endogenous variables
  - varlist_iv : instrument variables

IV estimator in Stata

- Options
  - e.g. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$,
    in which, x_1 is exogenous variables,
    x_2 is endogenous variables,
    and z_1, z_2 are instrument variables.

```
ivregress 2sls y x1 (x2 = z1 z2)
ivregress 2sls y x1 (x2 = z1 z2), r first

/* r--异方差稳健标准误
   first--report first-stage regression */
```

Subsection 3

An Example

IV estimator in Stata

- An Example
  - Mincer(1958) first researched the positive correlation between **salary** and **years of education**, but omitted the **ability** variable.
  - Griliches(1976) addressed the problem of omitted variable bias with IV method.
  - Data **:** Young Men's Cohort of the National Longitudinal Survey (NLS-Y).
  - Blackburn and Neumark(1992) Updated data of Griliches(1976).
  - Two-period panel data **:** the initial period is the earliest year in which the above variables have data; The end period is 1980.

IV estimator in Stata

- An Example : Data
  - lw (工资对数)
  - s (受教育年限)
  - age (年龄)
  - expr (工龄)
  - tenure (在现单位的工作年数)
  - iq (智商)
  - med (母亲的受教育年限)
  - kww (在“knowledge of the World of Work”测试中的成绩)
  - mrt (=1，已婚)
  - rns (=1，住在美国南方)
  - smsa (=1，住在大城市)
  - year (有数据的最早年份，1966-1973中的某一年)

IV estimator in Stata

1. Data Summary.

```
. use grilic, clear
. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
rns	758	.2691293	.4438001	0	1
rns80	758	.292876	.4553825	0	1
mrt	758	.5145119	.5001194	0	1
mrt80	758	.8984169	.3022988	0	1
smsa	758	.7044855	.456575	0	1
smsa80	758	.7124011	.452942	0	1
med	758	10.91029	2.74112	0	18
iq	758	103.8562	13.61867	54	145
kww	758	36.57388	7.302247	12	56
year	758	69.03166	2.631794	66	73
age	758	21.83509	2.981756	16	30
age80	758	33.01187	3.085504	28	38
s	758	13.40501	2.231828	9	18
s80	758	13.70712	2.214693	9	18
expr	758	1.735429	2.105542	0	11.444
expr80	758	11.39426	4.210745	.692	22.045
tenure	758	1.831135	1.67363	0	10
tenure80	758	7.362797	5.05024	0	22
lw	758	5.686739	.4289494	4.605	7.051
lw80	758	6.826555	.4099268	4.749	8.032

IV estimator in Stata

2. Test the correlation between ig(智商) and s(受教育年限).

```
. pwcorr iq s, star(.01)
             |       iq        s
-------------+------------------
          iq |   1.0000
           s |   0.5131*  1.0000

 * Significant positive correlation at the level of 1%,
 * correlation coefficient = 0.51.
```

IV estimator in Stata

3. Run an OLS regression.

- With robust standard error
- control variables : expr tenure rns smsa
- We are interested in s.

```
. reg lw s expr tenure rns smsa, r
Linear regression                               Number of obs     =        758
                                                F(5, 752)         =      84.05
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3521
                                                Root MSE          =     .34641
```

lw	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
s	.102643	.0062099	16.53	0.000	.0904523	.1148338
expr	.0381189	.0066144	5.76	0.000	.025134	.0511038
tenure	.0356146	.0079988	4.45	0.000	.0199118	.0513173
rns	-.0840797	.029533	-2.85	0.005	-.1420566	-.0261029
smsa	.1396666	.028056	4.98	0.000	.0845893	.194744
_cons	4.103675	.0876665	46.81	0.000	3.931575	4.275775

IV estimator in Stata

3. Run an OLS regression.

- Annual return on investment in education : 10.26% (Significant at the 1% level).
- **Overestimation** coefficient : omitted the **ability** variable.
- Ability is positively correlated with years of education.
- The contribution of ability to wages is included into the contribution of education.

IV estimator in Stata

4. Using iq(智商) as a **proxy variable** for ability, run an OLS regression.
   - Other **proxy variables**: High school test scores; Armed Forces Qualification Test(美国参军资格考试), etc.

```
. reg lw s iq expr tenure rns smsa, r
Linear regression                               Number of obs     =        758
                                                F(6, 751)         =      71.89
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3600
                                                Root MSE          =     .34454
```

lw	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
s	.0927874	.0069763	13.30	0.000	.0790921	.1064826
iq	.0032792	.0011321	2.90	0.004	.0010567	.0055016
expr	.0393443	.0066603	5.91	0.000	.0262692	.0524193
tenure	.034209	.0078957	4.33	0.000	.0187088	.0497092
rns	-.0745325	.0299772	-2.49	0.013	-.1333815	-.0156834
smsa	.1367369	.0277712	4.92	0.000	.0822186	.1912553
_cons	3.895172	.1159286	33.60	0.000	3.667589	4.122754

IV estimator in Stata

4. Using iq(智商) as a **proxy variable** for ability, run an OLS regression.
   - Annual return on investment in education reduced to 9.28%.
   - More reasonable, but still too large.
   - So iq(智商) is an **endogenous variable**.

IV estimator in Stata

5. Run 2SLS regression.

- Consider using med (母亲的受教育年限), kww (在"knowledge of the World of Work"测试中的成绩), mrt (=1，已婚), age (年龄) as IVs of iq(智商).
- With robust standard error

IV estimator in Stata

5. Run 2SLS regression.

```
Instrumental variables (2SLS) regression          Number of obs   =        758
                                                  Wald chi2(6)    =     355.73
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.2002
                                                  Root MSE        =     .38336
------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |  -.0115468   .0056376    -2.05   0.041    -.0225962   -.0004974
           s |   .1373477   .0174989     7.85   0.000     .1030506    .1716449
        expr |   .0338041   .0074844     4.52   0.000      .019135    .0484732
      tenure |    .040564   .0095848     4.23   0.000     .0217781     .05935
         rns |  -.1176984   .0359582    -3.27   0.001    -.1881751   -.0472216
        smsa |    .149983   .0322276     4.65   0.000     .0868182    .2131479
       _cons |   4.837875   .3799432    12.73   0.000       4.0932    5.58255
------------------------------------------------------------------------------
Instrumented:  iq
Instruments:   s expr tenure rns smsa med kww mrt age
```

IV estimator in Stata

5. Run 2SLS regression.

- Annual return on investment in education increased to 13.73%? (incredible)
- The contribution of iq(智商) to wages is negative? (incredible)
- We should check **instrument validity**.

IV estimator in Stata

6. Overidentification test.

- the number of instruments(4) $>$ the number of endogenous regressors(1)
- To test instrument exogeneity, thus overidentification test.

```
. estat overid
  Test of overidentifying restrictions:
  Score chi2(3)          =  51.5449  (p = 0.0000)
```

- Compute J-Statistic，some (or one) of the instrumental variables are invalid.
- We suspect mrt(=1，已婚), age(年龄) are invalid.
- Compute C-Statistic(检验部分工具变量不满足外生性) using *-ivreg2-*.

```
ssc install ivreg2
```

IV estimator in Stata

6. Overidentification test.

```
 /* ivreg2默认估计量为2SLS
    orthog(mrt age):检验(mrt,age)是否满足外生性 */

. ivreg2 lw s expr tenure rns smsa (iq = med kww mrt age), r orthog (mrt age)
IV (2SLS) estimation
--------------------

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity
                                                      Number of obs =       758
                                                      F(  6,   751) =     58.74
                                                      Prob > F      =    0.0000
Total (centered) SS     =  139.2861498                Centered R2   =    0.2002
Total (uncentered) SS   =  24652.24662                Uncentered R2 =    0.9955
Residual SS             =   111.39959                 Root MSE      =    .3834
------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |  -.0115468   .0056376    -2.05   0.041    -.0225962   -.0004974
           s |   .1373477   .0174989     7.85   0.000     .1030506    .1716449
        expr |   .0338041   .0074844     4.52   0.000      .019135    .0484732
      tenure |    .040564   .0095848     4.23   0.000     .0217781     .05935
         rns |  -.1176984   .0359582    -3.27   0.001    -.1881751   -.0472216
        smsa |    .149983   .0322276     4.65   0.000     .0868182    .2131479
       _cons |   4.837875   .3799432    12.73   0.000       4.0932     5.58255
------------------------------------------------------------------------------
```

IV estimator in Stata

6. Overidentification test.

```
. ivreg2 lw s expr tenure rns smsa (iq = med kww mrt age), r orthog (mrt age)
Underidentification test (Kleibergen-Paap rk LM statistic):              33.294
                                                    Chi-sq(4) P-val =    0.0000
------------------------------------------------------------------------------
Weak identification test (Cragg-Donald Wald F statistic):                10.538
                         (Kleibergen-Paap rk Wald F statistic):           9.585
Stock-Yogo weak ID test critical values:  5% maximal IV relative bias     16.85
                                         10% maximal IV relative bias     10.27
                                         20% maximal IV relative bias      6.71
                                         30% maximal IV relative bias      5.34
                                         10% maximal IV size              24.58
                                         15% maximal IV size              13.96
                                         20% maximal IV size              10.26
                                         25% maximal IV size               8.31
Source: Stock-Yogo (2005).  Reproduced by permission.
NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.
------------------------------------------------------------------------------
Hansen J statistic (overidentification test of all instruments):         51.545
                                                    Chi-sq(3) P-val =    0.0000
-orthog- option:
Hansen J statistic (eqn. excluding suspect orthog. conditions):           0.116
                                                    Chi-sq(1) P-val =    0.7333
C statistic (exogeneity/orthogonality of suspect instruments):           51.429
                                                    Chi-sq(2) P-val =    0.0000
Instruments tested:   mrt age
------------------------------------------------------------------------------
Instrumented:         iq
Included instruments: s expr tenure rns smsa
Excluded instruments: med kww mrt age
------------------------------------------------------------------------------
```

IV estimator in Stata

6. Overidentification test.

- 使用*-ivreg2-*得到的回归系数和稳健标准误与*-ivregress-*相同;
- 拒绝(mrt age)满足外生性的原假设;
- 考虑仅使用(med kww)作为iq的工具变量。

IV estimator in Stata

7. Run 2SLS regression again

```
. ivregress 2sls lw s expr tenure rns smsa (iq=med kww), r first
First-stage regressions
```

Number of obs	= 758
F(7, 750)	= 47.74
Prob > F	= 0.0000
R-squared	= 0.3066
Adj R-squared	= 0.3001
Root MSE	= 11.3931

iq	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
s	2.467021	.2327755	10.60	0.000	2.010052	2.92399
expr	-.4501353	.2391647	-1.88	0.060	-.9196471	.0193766
tenure	.2059531	.269562	0.76	0.445	-.3232327	.7351388
rns	-2.689831	.8921335	-3.02	0.003	-4.441207	-.938455
smsa	.2627416	.9465309	0.28	0.781	-1.595424	2.120907
med	.3470133	.1681356	2.06	0.039	.0169409	.6770857
kww	.3081811	.0646794	4.76	0.000	.1812068	.4351553
_cons	56.67122	3.076955	18.42	0.000	50.63075	62.71169

IV estimator in Stata

7. Run 2SLS regression again

```
Instrumental variables (2SLS) regression          Number of obs   =        758
                                                  Wald chi2(6)    =     370.04
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.2775
                                                  Root MSE        =     .36436
------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |   .0139284   .0060393     2.31   0.021     .0020916    .0257653
           s |   .0607803   .0189505     3.21   0.001      .023638    .0979227
        expr |   .0433237   .0074118     5.85   0.000     .0287968    .0578505
      tenure |   .0296442    .008317     3.56   0.000     .0133432    .0459452
         rns |  -.0435271   .0344779    -1.26   0.207    -.1111026    .0240483
        smsa |   .1272224   .0297414     4.28   0.000     .0689303    .1855146
       _cons |   3.218043   .3983683     8.08   0.000     2.437256    3.998831
------------------------------------------------------------------------------
Instrumented:  iq
Instruments:   s expr tenure rns smsa med kww
```

IV estimator in Stata

7. Run 2SLS regression again

- Annual return on investment in education reduced to 6.08%, which is reasonable.
- The contribution of iq(智商) to wages turns to positive again.

IV estimator in Stata

8. Check instrument validity
   - Check **IV relevance** : report the first stage.
     - Instrument perform well in the first stage.
     - A more formal test : F-statistic exceeds 10 (13.40), no Weak Instruments.

```
. estat firststage, all forcenonrobust
First-stage regression summary statistics
```

Variable	R-sq.	Adjusted R-sq.	Partial R-sq.	Robust F(2,750)	Prob > F
iq	0.3066	0.3001	0.0382	13.4028	0.0000

Shea´s partial R-squared

Variable	Shea´s Partial R-sq.	Shea´s Adj. Partial R-sq.
iq	0.0382	0.0305

```
Minimum eigenvalue statistic = 14.9058
Critical Values                          # of endogenous regressors:    1
Ho: Instruments are weak                 # of excluded instruments:     2
```

	5%	10%	20%	30%
2SLS relative bias		(not available)		

	10%	15%	20%	25%
2SLS Size of nominal 5% Wald test	19.93	11.59	8.75	7.25

IV estimator in Stata

8. Check instrument validity
   - Run and examine the **reduced form**.

```
qui reg lw s expr tenure rns smsa med, r
est store m1

qui reg lw s expr tenure rns smsa kww, r
est store m2

qui reg lw s expr tenure rns smsa med kww, r
est store m3

esttab m1 m2 m3,
    mtitle("reduced form:med" "reducedform:kww" "reducedform:med,kww")
    b(%6.3f) nogap compress
    star(* 0.1 ** 0.05 *** 0.01)
    ar2 order(med kww)
```

IV estimator in Stata

8. Check instrument validity
   - Run and examine the **reduced form**.

	(1) reduced_d	(2) reduced_w	(3) reduced_w
med	0.007		0.007
	(1.55)		(1.38)
kww		0.004**	0.004*
		(2.07)	(1.96)
s	0.100***	0.097***	0.095***
	(15.36)	(15.01)	(14.06)
expr	0.039***	0.037***	0.037***
	(5.88)	(5.42)	(5.52)
tenure	0.036***	0.032***	0.033***
	(4.49)	(4.04)	(4.10)
rns	-0.079***	-0.085***	-0.080***
	(-2.62)	(-2.87)	(-2.66)
smsa	0.138***	0.132***	0.131***
	(4.92)	(4.65)	(4.61)
_cons	4.058***	4.039***	4.002***
	(44.13)	(42.70)	(40.98)
N	758	758	758
adj. R-sq	0.349	0.351	0.352

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

IV estimator in Stata

8. Check instrument validity

- Run and examine the **reduced form**.
  - Reduced form is proportional to the casual effect of interest and is unbiased(OLS), so we should see the causal relation in the reduced form.
  - If you can't see the causal relation in the reduced form, it's probably not there.
  - **Notice!** Probably med is not exogenous enough (or not a very good IV).

IV estimator in Stata

8. Check instrument validity

- Check **exclusion restriction** : Overidentification test.
  - Instrumental variables (med kww) satisfy exogeneity.

```
. qui ivregress 2sls lw s expr tenure rns smsa (iq=med kww), r
. estat overid
  Test of overidentifying restrictions:
  Score chi2(1)          =  .151451  (p = 0.6972)
```

IV estimator in Stata

9. Check 2SLS estimates with LIML.

- LIML is less than precise than 2SIS but also less biased.
- If the results come out similar, be happy.

```
. ivregress liml lw s expr tenure rns smsa (iq=med kww), r
Instrumental variables (LIML) regression          Number of obs   =        758
                                                  Wald chi2(6)    =     369.62
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.2768
                                                  Root MSE        =     .36454
------------------------------------------------------------------------------
             |               Robust
          lw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          iq |   .0139764   .0060681     2.30   0.021     .0020831    .0258697
           s |   .0606362    .019034     3.19   0.001     .0233303    .0979421
        expr |   .0433416   .0074185     5.84   0.000     .0288016    .0578816
      tenure |   .0296237    .008323     3.56   0.000     .0133109    .0459364
         rns |  -.0433875    .034529    -1.26   0.209    -.1110631    .0242881
        smsa |   .1271796   .0297599     4.27   0.000     .0688512     .185508
       _cons |   3.214994   .4001492     8.03   0.000     2.430716    3.999272
------------------------------------------------------------------------------
Instrumented:  iq
Instruments:   s expr tenure rns smsa med kww
```

IV estimator in Stata

10. Put all estimates into one table.

```
qui reg lw s expr tenure rns smsa,r
est sto ols_noiq
qui reg lw iq s expr tenure rns smsa,r
est sto ols_iq
qui ivreg2 lw s expr tenure rns smsa (iq=med kww), r
est sto tsls
qui ivreg2 lw s expr tenure rns smsa (iq=med kww), r liml
est sto liml

esttab ols_noiq ols_iq tsls liml, mtitle    ///
    star(* 0.1 ** 0.05 *** 0.01)            ///
    b(%6.3f) nogap compress order(s iq)     ///
    stats(rkf j jp N r2_a, labels("First-stage F-statistic" ///
    "Overidentifying restrictions J-test and P-value" N r2_a) layout(@ `""@ (@)""´ @ @) )
```

IV estimator in Stata

10. Put all estimates into one table.

	(1) ols_noiq	(2) ols_iq	(3) tsls	(4) liml
s	0.103***	0.093***	0.061***	0.061***
	(16.53)	(13.30)	(3.21)	(3.19)
smsa	0.140***	0.137***	0.127***	0.127***
	(4.98)	(4.92)	(4.28)	(4.27)
iq		0.003***	0.014**	0.014**
		(2.90)	(2.31)	(2.30)
expr	0.038***	0.039***	0.043***	0.043***
	(5.76)	(5.91)	(5.85)	(5.84)
tenure	0.036***	0.034***	0.030***	0.030***
	(4.45)	(4.33)	(3.56)	(3.56)
rns	-0.084***	-0.075**	-0.044	-0.043
	(-2.85)	(-2.49)	(-1.26)	(-1.26)
_cons	4.104***	3.895***	3.218***	3.215***
	(46.81)	(33.60)	(8.08)	(8.03)
First-st~c			13.403	13.403
Overiden~t			0.151 (0.697)	0.151 (0.697)
N	758.000	758.000	758.000	758.000
r2_a	0.348	0.355	0.272	0.271

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

IV estimator in Stata

- Followed research
  - Return on investment in education is the core issue of labor economics.
  - Behrman et al(1980) compared **identical twins** with different years of education to controll the factors such as genetics and family background.
  - Angrist and Krueger(1991) used **the quarter of birth** as the instrumental variable of years of education.
  - Bound et al(1995) found the quarter of birth is a **weak** instrument.
  - Buckles and Hungerman(2012)'s latest research showed that the quarter of birth is **not** independent of family background.